HEALTH POLICY

Effects of Paying Physicians Based on their Relative Performance for Quality

Gary J. Young, JD, PhD^{1,2}, Mark Meterko, PhD^{1,2}, Howard Beckman, MD³, Errol Baker, PhD², Bert White, DMin, MBA¹, Karen M. Sautter, MPH¹, Robert Greene, MD³, Kathy Curtin, MBA, RN^{4,5}, Barbara G. Bokhour, PhD^{1,6}, Dan Berlowitz, MD, MPH^{1,6}, and James F. Burgess, Jr., PhD^{1,2}

¹Department of Health Policy and Management, Boston University School of Public Health, Talbot Building, 715 Albany Street, Boston, MA 02118-2526, USA; ²Department of Veterans Affairs, Center for Organization, Leadership and Management Research, Boston, MA, USA; ³Rochester Individual Practice Association, Rochester, NY, USA; ⁴Excellus Health Plans, Rochester, NY, USA; ⁵MedVantage, Inc., San Francisco, CA, USA; ⁶Center for Health Quality, Outcomes and Economic Research, Department of Veterans Affairs, Bedford, MA, USA.

BACKGROUND: Studies examining the effectiveness of pay-for-performance programs to improve quality of care primarily have been confined to bonus-type arrangements that reward providers for performance above a predetermined threshold. No studies to date have evaluated programs placing providers at financial risk for performance relative to other participants in the program.

OBJECTIVE: The objective of the study is to evaluate the impact of an incentive program conferring limited financial risk to primary care physicians.

PARTICIPANTS: There were 334 participating primary care physicians in Rochester, New York.

DESIGN: The design of the study is a retrospective cohort study using pre/post analysis.

MEASUREMENTS: The measurements adhere to 4 diabetes performance measures between 1999 and 2004.

RESULTS: While absolute performance levels increased across all measures immediately following implementation, there was no difference between the post- and pre-intervention trends indicating that the overall increase in performance was largely a result of secular trends. However, there was evidence of a modest 1-time improvement in physician adherence for eye examination that appeared attributable to the incentive program. For this measure, physicians improved their adherence rate on average by 7 percentage points in the year after implementation of the program.

CONCLUSIONS: This study demonstrates a modest effect in improving provider adherence to quality standards for a single measure of diabetes care during the early phase of a pay-for-performance program that placed physicians under limited financial risk. Further research is needed to determine the most effective

incentive structures for achieving substantial gains in quality of care.

KEY WORDS: pay-for-performance; financial incentives; quality of care; diabetes.

DOI: 10.1007/s11606-007-0185-5

© 2007 Society of General Internal Medicine 2007;22:872–876

INTRODUCTION

Across the United States, health plans and employer groups are adopting pay-for-performance programs that link financial incentives to quality of care. The basic concept is to present providers, typically physicians or hospitals, with financial incentives for achieving selected quality measures. There are currently more than 100 such programs operating in the United States, with many more under development¹. Furthermore, during the last few years, several bills have been introduced into Congress that would modify Medicare reimbursement policies to include financial incentives for quality performance^{2,3}.

Despite the rapid diffusion of the pay-for-performance incentive model, there is limited scientific support for the effectiveness of such programs. The few studies that do exist focus on programs with a similar type of incentive structure—those offering bonuses to providers for attaining predetermined performance targets^{4–8}. To date, these studies generally find, at best, modest effects on quality. However, while this type of incentive structure is common among pay-for-performance programs, the motivational effects of such incentive arrangements are uncertain for at least 2 reasons. First, providers whose performance historically has been above the predetermined performance target will likely have little motivation to improve further because the status quo is sufficient to obtain the bonus. Indeed, this pattern of behavior was observed in 1 recent study of a California-based program that offered bonuses to medical groups for attaining predetermined quality targets⁴. Second, the all-or-nothing proposition of predetermined performance targets may actually discourage some providers from trying to attain the target, particularly those whose past performance has been well below the level required for the bonus.

Received July 11, 2006

Revised November 27, 2006

Accepted March 15, 2007

Published online April 19, 2007

In this article, we report results from our investigation of a natural experiment with a pay-for-performance program that featured a very different incentive structure than previously studied programs. This program placed individual physicians at financial risk to receive rewards based on their performance for clinical quality relative to other physicians in the program. All participating physicians had opportunities to gain or lose compensation in direct proportion to their relative performance. Thus, the key question our study sought to address was whether an incentive program that places providers at financial risk for their relative performance on selected quality measures improves quality of care.

METHODS

The study focused on the pay-for-performance program of the Rochester (New York) Individual Practice Association (RIPA). Between 2000 and 2001, RIPA, in collaboration with the Excellus Health Plans, established the program as part of its capitated contract to care for approximately 300,000 Excellus Blue Choice health maintenance organization (HMO) patients. The program placed individual RIPA primary care physicians at limited financial risk for their performance with respect to clinical quality, patient satisfaction, and practice efficiency. The RIPA/Excellus incentive program provided a unique setting for investigating the effects of a single quality incentive program in that there were no competing pay-for-performance programs, either internally to RIPA or externally in the Rochester area. It is possible that some physician practices within RIPA may have had their own incentive schemes for physician productivity, but during the study period, the great majority of RIPA's physician practices were likely to be too small to have any such formal incentive arrangements.

RIPA is a multi-specialty physician network of over 3,500 providers, 600 of which are full-time primary care physicians involved in the pay-for-performance program. Most participating physicians practice in small or solo practices. Based on surveys RIPA has conducted of its affiliated physicians, Blue Choice patients accounted for between 35 and 55% of the patients for most of the affiliated physicians during the study period.

The financial risk element of the incentive program was linked to a withhold from physician fees. Each physician has had approximately 5% of his or her fees withheld to fund incentive pools. The money is distributed to physicians based on their relative performance in clinical quality, patient satisfaction, and practice efficiency. Since the program began, between \$12 and \$15 million has been distributed annually to participating providers. Each full-time primary care physician typically has contributed between \$8,000 and 14,000 of their annual income to an incentive pool and has been eligible to receive between 50 and 150% of their contribution depending on performance. For example, in 2003, the mean payout to RIPA Internists was 122 percent resulting in a possible return range of \$5,500 to \$16,500. In 2004, the mean Internist payout was 83% for a possible return range of \$4,500 to \$13,700.

The incentive program comprised several sets of quality performance measures selected by a RIPA/Excellus task force that relate to either chronic or acute care conditions. For purposes of the study, we focused on the set of performance measures for diabetes because complete data were available for this measure over a 6-year period that included baseline

and intervention periods. RIPA has required 5 diabetes performance measures annually: 2 Hemoglobin A1c tests, 1 lipoprotein density level (LDL) screening, 1 urinalysis/microalbumin, 1 flu vaccination, and 1 eye examination. Each of the 5 measures is a modification of a measure included in the Health Employer Data Information System (HEDIS). For example, whereas HEDIS specifications require 1 annual hemoglobin A1c test, RIPA/Excellus' incentive program requires 2 tests per year. To be eligible for this set of performance measures, a primary care physician (i.e., Internal or Family Medicine) must have been a RIPA physician for at least 24 months and have had 10 diabetic patients who were continuously enrolled in the Excellus commercial HMO plan (i.e., Blue Choice) and who had been under the care of that same physician for at least 24 months. RIPA sends physicians a report detailing their performance 3 times annually, as it did before establishing the incentive program.

For payment purposes, physicians were ranked annually according to their overall adherence score for each set of performance measures for which they qualify. Because there were 5 diabetes performance measures, a physician who had 10 diabetic patients enrolled in the Excellus commercial HMO plan would be accountable under the program for 50 services. If 25 services were provided, as determined through claims data, the physician would receive an adherence score of .50 (25/50). The actual payout to an individual physician was based on his or her weighted scores for clinical quality (40%), utilization (40%), and patient satisfaction (20%) with the actual dollar amount depending on the available funds in the incentive pools. At the time the program was introduced, the diabetes measures accounted for 50% of the total score for the quality component, which amounted to 20% of the total return under the pay-for-performance program (i.e., $.5 \times .4$). Given that physicians were required to contribute approximately 5% of their earnings to the incentive pool, the diabetes measures were worth approximately 1% of a typical primary care physicians' earnings (i.e., $.20 \times .05$) or \$1,500, assuming an annual income of \$150,000.

Study Design

Our investigation of the RIPA/Excellus program entailed a pre/posttest design focusing on 4 of the 5 diabetes performance measures. We excluded the performance measure for influenza vaccination because during 1 year of the study period, a vaccine shortage limited the ability of individual physicians to adhere to the measure. We used 6 years of annual performance data, from 1999 to 2004, that were obtained from a claim-based dataset that Excellus shared with RIPA. The unit of observation was the individual physician. The study was restricted to 334 primary care physicians who were members of RIPA for all 6 years of the study period and had qualified for the diabetes performance measure for at least 1 of the 6 years. For purposes of the analyses, we treated 2002 as the initial year of the intervention. While the quality-incentive program was announced in 2001, the first payout to physicians based on the diabetes measures was for their performance in 2002. Thus, we had 3 years of data for both the baseline and intervention periods. For each measure, we computed average annual physician adherence scores among all participating physicians.

Because the RIPA/Excellus incentive program was extended to all primary care physicians, we did not have a comparison group within the RIPA physician network itself. To account for

Table 1. Physician (n=334) adherence rates over time: mean (standard deviation)

Diabetes performance measures	Pre-intervention			Post-intervention		
	Yr1	Yr2	Yr3	Yr4	Yr5	Yr6
	1999	2000	2001	2002	2003	2004
	T1 (pre1)	T2 (pre2)	T3 (pre3)	T1 (post1)	T2 (post2)	T3 (post3)
HbA1c	.56 (.23)	.57 (.19)	.59 (.17)	.62 (.17)	.61 (.18)	.63 (.18)
Urinalysis	.61 (.21)	.63 (.20)	.66 (.16)	.67 (.18)	.69 (.17)	.70 (.17)
LDL	.58 (.24)	.64 (.19)	.69 (.16)	.74 (.15)	.75 (.15)	.79 (.13)
Eye Exam	.40 (.21)	.42 (.16)	.44 (.14)	.51 (.14)	.50 (.14)	.54 (.14)

secular trends in RIPA diabetes performance measures, we compared the pattern of RIPA scores to national and New York State trends in HEDIS scores over the study period. As noted, while several of the RIPA diabetes performance measures were modifications of HEDIS-specified measures, our comparison of RIPA scores to national and statewide scores focused on performance trends rather than levels. We obtained national HEDIS scores from the web site of the National Committee on Quality Assurance⁹ and New York State HEDIS scores through communication with the New York State Department of Health.

Data Analyses

Our analysis examined whether and to what degree the RIPA/Excellus quality incentive program was associated with improved physician adherence to the diabetes performance

measures. To address this question, we conducted a 2-way repeated measures analysis of variance (ANOVA) for each performance measure, which considered changes in both performance levels (pre-intervention vs post-intervention) and trends (T1 to T2 to T3 within pre- and post-intervention periods). In this analysis, statistically significant interactions between changes in performance levels and changes in trends indicate an actual change in performance beyond an extension of the pre-intervention pattern. In addition, we performed Neuman-Keuls *t* tests for post hoc multiple comparisons to assess on a year-to-year basis whether and where changes in performance scores for each measure were statistically significant throughout the study period. For this procedure, we compared differences in change scores for all pairs of adjacent years sequentially across the 6 years (e.g., pre2-pre1 vs pre3-pre2; post1-pre3 vs pre3-pre2).

Overview: Six Year Trends in Diabetes Care

Mean Percent of Patients Per Physician (n=334) Receiving Appropriate Care

Fig. 1. Overview: 6-year trends in diabetes care. Mean percent of patients per physician (n=334) receiving appropriate care

Table 2. Impact of RIPA quality targets and incentives on care of diabetics 1999–2004: post hoc comparisons

Diabetes pathway component	Change in adherence rate		Difference in rate of change (pre3–pre2) vs (post1–pre3)
	Pre2 to pre3	Pre3 to post1	
HbA1c	.0176	.0262*	.0086
Urinalysis	.0278*	.0105	.0172
LDL	.0439‡	.0472‡	.0033
Eye	.0195	.0713‡	.0518‡

* $P \leq .05$ † $P \leq .01$ ‡ $P \leq .001$

RESULTS

The average number of diabetic patients per physician over the 6-year period was 22.3 (SD 13.7). Table 1 presents basic descriptive statistics for each of the performance measures for the 3 pre-intervention and 3 post-intervention years that were analyzed using the repeated measures analysis of variance (ANOVA). The means over time are plotted in Figure 1.

The results of the repeated measures ANOVA indicate that for each of the performance measures, there was a statistically significant increase in performance levels after the introduction of the program (significant pre–post main effect from repeated measures ANOVA) with the largest increases being seen for LDL screening and eye examination. However, based on the absence of a significant interaction term for each

measure in this context, the post-intervention trends were not different from the pre-intervention trends, indicating that the overall pattern of performance did not change after program introduction. The Neuman–Keuls post hoc t tests (see Table 2), which provides a more fine-grained analysis for assessing year-to-year changes in performance scores, revealed a significant difference in the rate of improvement for only 1 of the 4 measures during the year after the intervention (pre 3 to post 1) compared to the year preceding the intervention (pre 2 to pre 3). Specifically, there was a significant increase of 7 percentage points in the eye examination performance score for the year after the introduction of the program, although this rate of change did not persist in the 2 subsequent years of the post-intervention period. Figure 2 illustrates this finding with comparative data for the nonsignificant post-intervention change for HbA1c compliance. In addition, the increase in the RIPA eye examination score was contrary to the trends observed in the HEDIS eye examination scores nationally and statewide during this period, which were largely flat from 2000 to 2002. Thus, the observed increase did not appear to be attributable to a secular trend. These results did not change when we accounted statistically for physicians' primary care specialty (Internist vs Family Medicine) and practice arrangement (solo vs group).

DISCUSSION

We investigated the effects of a program that placed each physician at some financial risk for their quality performance

**Change in Adherence Rates Over Time:
HbA1c and Eye Exam**

Fig. 2. Change in adherence rates over time: HbA1c and eye exam

relative to other physicians in the program. While all performance measures improved after the program was implemented, only the eye examination measure showed a statistically significant change, and even this effect appeared to be a 1-time improvement in the score. It is not readily apparent why improvement in the eye examination measure exceeded that of the other performance measures, although it is noteworthy that the eye examination score was the lowest of the 4 performance measures at baseline, and thus, there was more opportunity for improvement.

Our results raise the question of why such a program might not have had a greater impact than the one observed. A number of possibilities deserve consideration. One consideration is the time frame required for physicians to respond to financial incentives and adopt new practice methods. Our study was confined to the effects of the RIPA program within 3 years of its implementation. Given the complexity of changing physician's knowledge of and response to new reimbursement policies, evaluations of the effectiveness of quality incentive programs may require longer timeframes than those used for the present study.

Another consideration is the way physicians are organized in the Rochester area. As noted, most RIPA physicians work in solo or small practices. Although the financial incentive may have motivated them to focus their time and energy to achieving the program's performance measures, most physicians who work in small practices lack the infrastructure that health care experts often say is important to improving quality of care¹⁰. The lack of such infrastructure such as electronic medical records, may have limited the effects of the incentive program, although we did not investigate this question specifically. It is possible that incentive programs aimed at very large medical groups, hospitals, or other large health care delivery organizations that have the resources to invest in clinical infrastructure may experience larger effects from quality incentive programs than what was observed in this study. However, the choice of rewarding individual physicians directly versus medical groups is itself a potentially controversial design consideration for pay-for-performance programs. Indeed, there is evidence that when medical groups serve as a financial intermediary for pay-for-performance programs, they distribute reward money in ways that may attenuate the financial incentives for individual physicians¹¹.

The results of the study also raise the question of whether the financial incentive was large enough to motivate physicians to change their clinical practice. For most physicians in the study, the potential payout they could receive represented approximately between 5 and 10% of their practice income. While some experts consider this amount to be of sufficient size¹, no firm evidence exists to support this figure. Furthermore, the diabetes measures accounted for no more than 20% of the total payout to physicians included in the study. As the number of measures in a pay-for-performance program increases, the return for improving any 1 measure decreases, and this may limit the program's potential to improve quality of care. In designing such programs, careful consideration may need to be given to the number of measures selected and each measure's contribution to the total payout in relation to what the measure requires in time and effort from providers.

The study has 2 notable limitations. First, because the study was confined to physicians practicing in the Rochester

area, the results may not generalize to other settings. Second, by capitalizing on a natural experiment, the study lacked the methodological rigor of a randomized trial. While some commentators have called for greater use of randomized trials for evaluating pay-for-performance programs^{12,13}, whether this methodology can be used both efficiently and effectively to evaluate pay-for-performance programs remains to be seen given the array of program design options that are available and require testing.

In conclusion, our study provides the first systematic investigation of a pay-for-performance program that placed providers at financial risk for their relative quality performance. As interest in pay-for-performance continues to grow, we may well see innovations in the design and implementation of quality incentive models. Such innovations will provide important opportunities to evaluate which types of designs appear to work best in the pursuit of improving quality of care.

Acknowledgements: We gratefully acknowledge the contribution by Greg Partridge in facilitating access to the data we analyzed in our study. We thank Matthew Guldin for his research assistance. This study was supported by grants from the Agency for Healthcare Research and Quality and the Robert Wood Johnson Foundation.

Conflict of Interests Statement: None Disclosed.

Corresponding Author: Gary J. Young, JD, PhD; Department of Health Policy and Management, Boston University School of Public Health, Talbot Building, 715 Albany Street, Boston, MA 02118-2526, USA; (e-mail: health@bu.edu).

REFERENCES

1. Baker G, Carter B. Provider pay-for-performance incentive programs: 2004 national study results. San Francisco: Med-Vantage, Inc; 2005.
2. Medicare Value-Based Purchasing for Physicians' Services Act of 2005, H.R. 3617, 109th Cong., 1st Sess. (2005).
3. Medicare Value Purchasing Act of 2005, S. 1356, 109th Cong., 1st Sess. (2005).
4. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA*. 2005;294:1788-93.
5. Amundson G, Solberg LI, Reed M, Martini EM, Carlson R. Paying for quality improvement: compliance with tobacco cessation guidelines. *Jt Comm J Qual Saf*. 2003;29:59-65.
6. Roski J, Gregory R. Performance measurement for ambulatory care: moving towards a new agenda. *Int J Qual Health Care*. 2001;13:447-53.
7. Kouides RW, Bennett NM, Lewis B, Cappuccio JD, Barker WH, LaForce FM. Performance-based physician reimbursement and influenza immunization rates in the elderly: the primary-care physicians of Monroe County. *Am J Prev Med*. 1998;14:89-95.
8. Hillman AL, Pauly MV, Kerstein JJ. How do financial incentives affect physicians' clinical decisions and the financial performance of health maintenance organizations? *N Engl J Med*. 1989;321:86-92.
9. The state of health care quality. Washington, D.C.: National Committee for Quality Assurance, 2005 and 2003. (Accessed November 4, 2005 at http://www.ncqa.org/Docs/SOHCQ_2005.pdf and <http://www.ncqa.org/Communications/Publications/03annual.pdf>.)
10. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282:1458-65.
11. Bokhour BG, Burgess JF, Hook JM, et al. Incentive implementation in physician practices: a qualitative study of practice executive perspectives on pay for performance. *Med Care Res Rev*. 2006;63:73S-95S.
12. Dudley RA. Pay-for-performance research: how to learn what clinicians and policy makers need to know. *JAMA*. 2005;294:1821-3.
13. Chassin MR. Does paying for performance improve the quality of healthcare? *Med Care Res Rev*. 2006;63:122S-55S.